Silky shark abundance index based on CPUE standardization of French Indian Ocean tropical tuna purse seine observer bycatch data

David M. Kaplan[1,2,*], Mariana Travassos Tolotti[1,2]

Abstract

Here we present an annual abundance index for silky shark for the period 2012 to 2021 based on standardized silky shark catch per FOB fishing set for the French tropical tuna purse seine fishery in the Indian Ocean. The methodology used general additive models with mixed effects (GAMMs) in a novel approach using three submodels, including one model for tropical tuna catch per set, and two others that form the components of a Δ log-normal model (i.e., a presence-absence model and an abundance when present model) for silky shark catch per set with total tuna catch used as predictors in those models. Results indicate an overall increasing trend in silky shark abundance over the study time period, though when interannual variability in catch of target tunas is removed model CPUE predictions are found to be more or less stable over time. Though there is evidence to support a potential increasing trend in abundance, there are a number of reasons to believe that these predictions may be overly optimistic, including low data coverage and a biased spatial distribution of fishing effort and observer coverage of bycatch in the original two years of the data (2012-2013) and poorly understood impacts of the implementation of a quota on yellowfin tuna catch in 2017.

Keywords

prediction intervals; general additive model (GAM); silky shark (*Carcharhinus falciformis*); catch per unit effort (CPUE)

1 Introduction

The status of bycatch species in general, and silky shark (*Carcharhinus falciformis*) in particular, are often difficult to assess in large part due to a lack of robust abundance indices that can be used to drive stock status models (Rice & Harley 2013, IOTC Secretariat 2022a). In the Indian Ocean, silky shark status was most recently classified as "unknown" due to uncertainty in the relationship between longline CPUEs and abundance (IOTC Secretariat 2022a). In this context, it is advisable to examine alternative CPUE indices from other fisheries.

Catch of tropical tunas by purse seine vessels regularly includes bycatch of silky sharks (Kaplan et al. 2014, Forget et al. 2021, Dumont et al. submitted). In the Indian Ocean, purse seine silky shark bycatch almost exclusively occurs in sets on schools associated with floating objects (hereafter referred to as "FOB sets"). Though observer data on purse seine bycatch has existed at low coverage levels for many years, observer coverage drastically increased starting in 2013-2014 and now covers more than a third of all sets by the French purse seine fleet in the Indian Ocean (IOTC Secretariat 2022b). Though these data have been used extensively, e.g., to estimate bycatch levels and species composition (Amandè et al. 2012, Torres-Irineo et al. 2014), estimate the impacts of spatial management (Kaplan et al. 2014) and predict spatial distributions of bycatch species (Mannocci et al. 2020), to date these data have not to our knowledge been used to estimate abundance indices for silky sharks. This absence has in part been due to the short time series of observer bycatch data with high coverage levels, but this limitation is gradually changing as time series become longer and we now have approximately a decade of observer bycatch data.

[1]MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France
[2]Institut de Recherche pour le Développement (IRD), av. Jean Monnet, CS3017, 34203 Sète Cédex, France

[*]Corresponding author. Email: david.kaplan@ird.fr

Here, we develop a silky shark abundance index based on (by)catch per unit effort (CPUE) standardization of catch and bycatch data from the French tropical tuna purse seine fishery over the period 2012 - 2021. There are at least two major complications to standardizing silky shark CPUEs (Dumont et al. submitted): (1) it is known that catch of target tunas impacts bycatch of silky sharks, so this predictor must be integrated into standardization efforts; (2) the number of zero catches found among purse seine fishing sets is high, at level inconsistent with the use of the simplest distribution families for residuals. To address the first impediment, we first develop a model for total catch of target tunas, which is then used as an input to models for silky shark bycatch when making predictions for CPUE standardization. The second impediment is addressed by using a Δ approach to estimating silky shark catch per FOB set wherein presence-absence of silky sharks and the number of silky shark caught in sets for which silky sharks are known to be present are estimated in two separate models, predictions from which are multiplied together to make a final prediction of silky shark bycatch per set.

2 Material and methods

2.1 Fishing activity data

The primary data used in this study are catch data for target tunas from logbooks (i.e., mandatory declarations) and for bycatch species from observer data from the Indian Ocean French tropical tuna purse seine fishery between 2012 and 2021. The Exploited Tropical Pelagic Ecosystems Observatory (Ob7) of the French National Research Institute for Sustainable Development (IRD) collects and manages fishery and observer data from the French tropical tuna purse seine fleet in the Atlantic and Indian Oceans. Though the approach we develop could be applied in the Atlantic and elsewhere, we examine here data from the Indian Ocean for which 33.6% of FOB sets are covered by observer data over the study time period. Data on fishing set vessel, time, location and catch in tonnes of the three major target tunas (skipjack tuna, yellowfin tuna and bigeye tuna) were obtained from captain logbooks data corrected for bias in species composition and total catch based on port sampling using the T3 procedure (Pianet et al. 2000). Data on bycatch per species in number of individuals for each fishing set were obtained from onboard observers as part of two PS observer programs: the European Union-funded "Data Collection Framework" (DCF; Reg 2017/1004 and 2016/1251) and the "Observateur Commun Unique Permanent" (OCUP) program initiated and funded by the PS fishing industry (Goujon et al. 2017). Each line of the data corresponds to a fishing set, i.e., a single fishing activity involving encircling a tuna school with the seine and bringing the fish caught within the seine on board (Table 1).

The following filters were applied to catch-effort data before including them in CPUE standardization models:

1) Only logbook entries corresponding to single non-null FOB fishing sets were considered. In particular, a small percentage of logbook entries corresponding to multiple fishing sets were eliminated as including these data is both statistically complicated and could induce errors (e.g., in individual set position or catch).
2) Only sets corresponding to French-flagged vessels were included (i.e., data from French vessels flagged under other nations were not used).
3) A small number (31) of sets with zero catch of tropical tunas (i.e., skipjack, bigeye and yellowfin) were eliminated as these generally correspond to catch of temperate tunas (e.g., albacore) or other less common species, all of which are rare and generally occur well outside the core study area of the western-central Indian Ocean.
4) Whether a vessel was fishing on a FOB equipped with one of its own tracking buoys and the approximate set time relative to local sunrise were calculated based on comparisons of logbook, T3-corrected, VMS and dFAD tracking buoy data. When these two predictor variables could not be calculated (e.g., due to a lack of correspondence between databases or unavailability of VMS data), sets were not included in analyses.

For running models of silky shark bycatch, data were limited to sets with observer coverage. Due to differences in unique identifiers used in logbook and observer databases, a matching procedure was needed to associate catch-effort data with observer data. Sets in the two databases were considered to be the same if the following matching conditions were met:

1) Fishing vessel and date match
2) The total number of fishing sets carried out by the vessel on the given day match

3) The sets in the two database occur in the same order (i.e., the first logbook set is matched with the first observer set, the second with the second, etc.).
4) The declared school types (i.e., FOB set or free-swimming school set) match in the two databases
5) The spatial distance between the declared geographical positions of the sets in the two databases was less than 20 km.

The final data consisted of 15967 FOB sets, of which 5357 have observer data. When counting the number of silky sharks caught, all observed individuals were included, including those that were live released.

2.2 Modeling bycatch per set with a Δ method

The model of silky shark catch per FOB set is developed using three sub-models: (1) a model of tropical tuna catch per set; (2) a model of presence-absence of silky shark bycatch; and (3) a model of the number of silky shark individuals caught in a set given that the number is greater than zero (i.e., limiting data to sets for which silky shark were present). All three models were estimated using general additive models with mixed effects (GAMMs), the first and last assuming normally distributed residuals in the log of the catch and the second assuming a binomial distribution. In all cases, a unique vessel identifier was treated as a categorical random effect in models.

Other predictor variables consist of spatial variables longitude and latitude, temporal variables year and quarter, fishing vessel capacity, the number of hours since sunrise that the set was estimated to occur at, and whether or not the set was estimated to occur on a FOB being actively followed by the fishing vessel via its tracking buoy (Table 2). Fishing vessel capacity has often been included in CPUE standardization models (Kaplan et al. 2023), but its potential effects on bycatch rates are less clear. Time relative to sunrise has been shown to both effect the composition of species around floating objects (Forget et al. 2015) and thereby impact catch (Kaplan et al. 2023). Whether or not vessels are fishing on FOBs equipped with their own echosounder-equipped tracking buoys has been shown to impact catches of target species (Wain et al. 2021; note that by 2012 the majority of tracking buoys used by the French fleet were echosounder equipped and by 2014 100% were echosounder equipped).

The specific model used for estimating tropical tuna catch was:

```
## logcatch ~ te(ecd_lon, ecd_lat, by = year_fact, k = 7) + s(capacity,
##     k = 6) + s(hours_since_sunrise, k = 6) + follow + year_fact *
##     quarter_fact
```

with `vessel_id` included as a random effect. `logcatch` is the natural logarithm of the total catch of target tunas in a set, `kcap`=8, and `year_fact` and `quarter_fact` are the same as the `year` and `quarter` variables, but transformed from integer into factor variables (in the parlance of R). Note that these models have been constructed so that there is a detailed interaction between space and year, but seasonal effects are only incorporated view the `quarter_fact` categorical variable and indirectly via spatial parameters. This choice was made because initial models including a space-quarter interaction, but no space-year interaction, had a temporal trend in residuals suggesting that a space-year interaction was needed, and because models including both a space-quarter and a space-year interaction failed to converge in a timely fashion. Models including the space-year interaction only converged in a reasonable amount of time if the `k` parameters in smooths were adjusted downward (as shown in the equation above), particularly in the space-year term.

The specific model used for estimating presence-absence of silky sharks was:

```
## silky_pres ~ s(catch, k = 7) + te(ecd_lon, ecd_lat, by = year_fact,
##     k = 7) + s(capacity, k = 6) + s(hours_since_sunrise, k = 6) +
##     follow + year_fact + quarter_fact
```

where `silky_pres` is a Boolean variable indicating if silky sharks were caught in the fishing set (with observer coverage). Other than the inclusion of the smooth on `catch`, this model differs from the model for target tropical tuna catch in that it includes no interaction between `year_fact` and `quarter_fact`. This was done due to a lack of observer data for some year-quarter combinations in the initial two years of the study time period.

The specific model used for estimating the number of silky sharks caught in a set given that at least one was caught was:

```
## logsilky ~ s(catch, k = 7) + te(ecd_lon, ecd_lat, by = year_fact,
##     k = 7) + s(capacity, k = 6) + s(hours_since_sunrise, k = 6) +
##     follow + year_fact + quarter_fact
```

where `logsilky` is the natural logarithm of the number of silky sharks caught.

As it is a bit unusual to use (log)catch as the predicted variable in one model and a predictor in another model where all models share other predictors (thereby, putting into question the independence of tuna catch as a predictor for silky shark catch relative to other predictors, such as set time), we also did a set of model runs for silky shark presence-absence and abundance when present where `catch` was replaced by either residuals or log-residuals from the model for catch of tropical tunas. As these models had very similar diagnostics and deviance explained to those of the original models using `catch`, likely due to the overall limited variance explained by the models, we decided not to pursue these further and use the original models for further predictions.

2.2.1 Standardized CPUE predictions

Our main objective with the CPUE standardization was to standardize fishing effort (i.e., the number of fishing sets) seasonally and over space. To do this, we first determined the average number of fishing sets per year in each $1° \times 1°$-quarter strata occurring in our data. These numbers were then rounded to the nearest integers (Fig. 1), giving us the number of fishing sets to simulate in each $1° \times 1°$-quarter strata when predicting standardized CPUEs. Predictions for the given number of sets in each strata were averaged over each year to yield an average number of silky sharks caught per FOB set, which was used as our final silky shark abundance index. This had the effect of eliminated from the prediction grid all outlying areas with <0.5 FOB fishing sets per strata on average (Fig. 1).

This prediction spatial grid varied by quarter, but was repeated identically over years to standardize the number and spatio-seasonal distribution of fishing effort across years. Additional predictors `hours_since_sunrise` and `capacity` were fixed at their median values found in Table 2, and `follow` was arbitrarily set to `TRUE`. When predicting silky shark bycatch, the tuna catch predictor was determined as the output predictions from the model for tuna catch, as described below.

In order to estimate an average number of silky sharks per FOB set and its uncertainty, for each fishing set in the prediction grid (of which there may be multiple per strata, depending on the average number of FOB fishing sets in a strata), we estimated a prediction interval for the silky shark catch in the set. Prediction intervals differ from the more common confidence intervals in that while the latter estimates the certainty in the expectation value (i.e., the mean), the prior estimates the uncertainty in new predictions (in this case potential silky shark catch in a fishing sets. Prediction intervals can be evaluated directly for general linear models (GLMs) using functionality already integrated into standard R functions for GLMs. However, for GAMMs, standard software tools for calculating prediction intervals are lacking. Instead, we decided to calculate prediction intervals using a bootstrap approach following Andersen (2022).

For each of the three submodels, the basic procedure was the same, but the details differ somewhat. We first describe the procedure for the GAMM for catch of target tunas and then proceed to models for silky shark catch. First, we randomly generated 300 sets of potential coefficients of the GAMM for catch of target tunas. This was done using the Cholesky Trick following the approach of Andersen (2022). Simply put, this approach uses the optimal coefficients estimated by the model and the standard errors of these coefficients to estimate a distribution for each coefficient from which the potential coefficients are drawn.

Then, for each of these 300 potential model coefficients, model predictions in link-function space including uncertainty in parameters were calculated as the linear sum of the coefficients multiplied by the corresponding predictors. These were then back-transformed into response units using the appropriate inverse link function (in the case of normally distributed data, this would be the identity function, whereas for a binomial model this would be the inverse logit function). Final predictions were generated by randomly drawing from the appropriate model distribution family (i.e., normal distribution for the tuna catch and silky shark abundance models, binomial for the silky shark presence-absence model), including for two parameter distributions, such as the normal distribution used for modeling tuna catch, the residual (i.e., unexplained) variance of each of GAMM (i.e., the scale parameter). This produced 300 simulated model predictions of log tuna catch for each fishing set in the grid, with exponentiation being used to obtain the final predictions for tuna catch. If our ultimate goal was estimating tuna catch, then quantiles of the 300 simulations could be used to estimate the prediction uncertainty in the catch of each set.

Calculating prediction intervals for the silky shark presence-absence and log-abundance when present GAMMs proceeded in an identical fashion with the exception that the 300 simulated tuna catch values for each fishing set were used as the `catch` predictor values when making predictions from these models. This procedure produced 300 simulations of silky shark presence-absence and log-abundance for each fishing set in the prediction grid. These 300 simulations incorporate spatio-temporal variability in catch per set, uncertainty in catch per set and unexplained variance in the number of silky sharks caught per set. For each of the 300 simulations, presence-absence and the exponential of log-abundance were multiplied together. To obtain the final average number of silky sharks per set, predictions were summed over years for each of the 300 simulations and the distribution of the 300 simulations were used to estimate the mean, median and 95% prediction interval for average number of silky sharks per set for each year.

As there is some uncertainty as to whether variability across years in tuna catch per set is due to true changes in abundance as opposed to changes in catchability (e.g., due to technology), a second abundance index was calculated using the procedure described above, but with all interannual variability in catch per set (and the variability across the 300 simulations) removed when predicting silky shark presence-absence and log-abundance. Catch per set was allowed to vary by quarter and over space, but values were the same for all years and all simulations within a given quarter and spatial stratum.

A final set of simulations was performed using, instead of the prediction grid described above, the original catch-bycatch data used to fit models as the grid for making predictions. The objective of these simulations was to produce a standardized silky shark catch per set that should be directly comparable to the nominal silky shark catch per set (i.e., as a check of the model fit to the data).

2.3 Numerical tools

All analyses were carried out using R version 4.3.1 (2023-06-16) (R Core Team 2023) and the `tidyverse` package version 2.0.0 (Wickham et al. 2019). Models were fit using the R package `mgcv` version 1.9.0 (Wood 2017) and data visualizations used the packages `ggplot2` version 3.4.3 (Wickham 2016), `mgcViz` version 0.1.9 (Fasiolo et al. 2018) and `tmap` version 3.3.4 (Tennekes 2018).

3 Results

3.1 Variability across years in fishing effort and observer coverage

Both the number of FOB fishing sets in our data and the observer coverage of those sets varied in a non-random way across years, with relatively low coverage in 2012-2013 and much higher coverage from 2014 until 2019, after which time the COVID pandemic (temporarily) reduced coverage (Fig. 2). In particular, the number of FOB fishing sets with observer coverage is on the order of 100 for 2012-2013, indicating that relatively little data is available for accurate model estimation for those years. As such, results for these years should be treated with care.

The nominal fraction of fishing sets for which silky sharks were present is relatively more stable over years, with values typically around 80% for most years (Fig. 2b). Nevertheless, there does appear to be a slight increasing trend in prevalence over years, with values typically in the 65-75% range in the first five years of the data and then increasing to 80-90% in the last five years of the data.

3.2 Model summaries, diagnostic plots and marginal effect plots

GAMMs are actually implemented as combination of a linear mixed-effects (LME) model to estimate the random effects (of vessel identifier in our case) and a GAM to estimate the non-random effects after removing the random effects. Below I present basic diagnostic summaries and figures for each of the three submodels in the following order:

1) An ANOVA table of the LME model
2) A standard residuals versus fitted plot for the LME model
3) Summary output from the GAM model
4) A QQ-plot for the GAM model
5) Marginal effects plots for the single smoothed effects for `hours_since_sunrise` and `capacity` and `catch` where appropriate

I have not included 2D marginal effects plots for the spatial lon,lat smooths because there are multiple of these and I have found them typically to be difficult to adequately visualize and interpret.

3.2.1 GAMM for log-catch of target tunas

The ANOVA table for the LME model is:

```
##   numDF denDF  F-value p-value
## X    73 15880 1231.752  <.0001
```

This indicates that the effect of `vessel_id` is significant in this model and the other two models. Nevertheless, the variance explained by this effect (not shown) is small.

The residuals versus fitted for the LME model can be found in Fig. 3.

The summary output for the GAM component of the GAMM model is as follows:

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## logcatch ~ te(ecd_lon, ecd_lat, by = year_fact, k = 7) + s(capacity,
##     k = 6) + s(hours_since_sunrise, k = 6) + follow + year_fact *
##     quarter_fact
##
## Parametric coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2.8284684  0.1061157  26.655  < 2e-16 ***
## followTRUE                    0.0174224  0.0152968   1.139 0.254737
## year_fact2013                 0.0008102  0.1368729   0.006 0.995277
## year_fact2014                -0.0626388  0.1220547  -0.513 0.607817
## year_fact2015                -0.0985015  0.1359484  -0.725 0.468739
## year_fact2016                -0.1620337  0.1189769  -1.362 0.173251
## year_fact2017                -0.0033275  0.1181682  -0.028 0.977536
## year_fact2018                 0.3924396  0.1150013   3.412 0.000645 ***
## year_fact2019                 0.2616534  0.1158788   2.258 0.023960 *
## year_fact2020                -0.1716862  0.1174487  -1.462 0.143817
## year_fact2021                 0.3143926  0.1223330   2.570 0.010180 *
## quarter_fact2                -0.2485444  0.1197607  -2.075 0.037971 *
## quarter_fact3                 0.1041333  0.1052935   0.989 0.322687
## quarter_fact4                -0.1355606  0.1125279  -1.205 0.228343
## year_fact2013:quarter_fact2   0.0647511  0.1532341   0.423 0.672620
## year_fact2014:quarter_fact2  -0.0537038  0.1476370  -0.364 0.716045
## year_fact2015:quarter_fact2  -0.2151901  0.1587569  -1.355 0.175287
## year_fact2016:quarter_fact2   0.3714564  0.1383632   2.685 0.007268 **
## year_fact2017:quarter_fact2   0.1429372  0.1375656   1.039 0.298798
## year_fact2018:quarter_fact2  -0.1491014  0.1349955  -1.104 0.269397
## year_fact2019:quarter_fact2  -0.2298647  0.1433905  -1.603 0.108940
## year_fact2020:quarter_fact2   0.0405350  0.1392535   0.291 0.770988
## year_fact2021:quarter_fact2  -0.1839844  0.1398894  -1.315 0.188457
## year_fact2013:quarter_fact3  -0.1282132  0.1564259  -0.820 0.412433
## year_fact2014:quarter_fact3  -0.0158381  0.1280426  -0.124 0.901559
## year_fact2015:quarter_fact3   0.0711882  0.1440590   0.494 0.621200
## year_fact2016:quarter_fact3   0.2683465  0.1275038   2.105 0.035340 *
## year_fact2017:quarter_fact3   0.0170349  0.1293554   0.132 0.895231
## year_fact2018:quarter_fact3  -0.0662392  0.1217022  -0.544 0.586262
## year_fact2019:quarter_fact3  -0.0287928  0.1259627  -0.229 0.819197
## year_fact2020:quarter_fact3   0.5065893  0.1309300   3.869 0.000110 ***
## year_fact2021:quarter_fact3  -0.1345599  0.1292726  -1.041 0.297938
## year_fact2013:quarter_fact4   0.1769253  0.1624627   1.089 0.276161
## year_fact2014:quarter_fact4  -0.0749964  0.1365623  -0.549 0.582894
```

```
## year_fact2015:quarter_fact4  0.1034049  0.1440962   0.718 0.473008
## year_fact2016:quarter_fact4  0.3025594  0.1296768   2.333 0.019651 *
## year_fact2017:quarter_fact4  0.4186949  0.1313726   3.187 0.001440 **
## year_fact2018:quarter_fact4 -0.1575506  0.1290238  -1.221 0.222067
## year_fact2019:quarter_fact4  0.0164347  0.1319928   0.125 0.900911
## year_fact2020:quarter_fact4  0.4251032  0.1341127   3.170 0.001529 **
## year_fact2021:quarter_fact4  0.2195613  0.1335403   1.644 0.100164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                        edf Ref.df       F  p-value
## te(ecd_lon,ecd_lat):year_fact2012  8.268  8.268   3.102  0.00132 **
## te(ecd_lon,ecd_lat):year_fact2013  7.443  7.443   3.781  0.00108 **
## te(ecd_lon,ecd_lat):year_fact2014  6.951  6.951   7.385  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2015 12.742 12.742   4.009  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2016  6.398  6.398   4.602 7.81e-05 ***
## te(ecd_lon,ecd_lat):year_fact2017  9.129  9.129   5.057  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2018 12.622 12.622   5.094  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2019  7.402  7.402   1.688  0.10611
## te(ecd_lon,ecd_lat):year_fact2020  9.734  9.734   3.082  0.00125 **
## te(ecd_lon,ecd_lat):year_fact2021 11.622 11.622   3.403 5.94e-05 ***
## s(capacity)                        3.079  3.079  14.178  < 2e-16 ***
## s(hours_since_sunrise)             2.677  2.677 179.762  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0972
##   Scale est. = 0.8424    n = 15967
```

A QQ-plot for the GAM component of the GAMM model for catch of target tunas can be found in Fig. 4. Marginal effects plots for smoothed predictors in the GAM model can be found in Fig. 5.

3.2.2 GAMM for presence-absence of silky shark

The ANOVA table for the LME model is:

```
##   numDF denDF  F-value p-value
## X    47  5299 5.266348  <.0001
```

The residuals versus fitted for the LME model can be found in Fig. 6.

The summary output for the GAM component of the GAMM model is as follows:

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## silky_pres ~ s(catch, k = 7) + te(ecd_lon, ecd_lat, by = year_fact,
##     k = 7) + s(capacity, k = 6) + s(hours_since_sunrise, k = 6) +
##     follow + year_fact + quarter_fact
##
## Parametric coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.110105   0.634681   1.749   0.0803 .
## followTRUE     0.136072   0.076276   1.784   0.0745 .
## year_fact2013  0.008631   0.669391   0.013   0.9897
## year_fact2014  0.177431   0.578654   0.307   0.7591
## year_fact2015  0.201571   0.581902   0.346   0.7291
## year_fact2016  0.150189   0.575474   0.261   0.7941
```

```
## year_fact2017  0.804170   0.579078   1.389   0.1650
## year_fact2018  0.702205   0.577700   1.216   0.2242
## year_fact2019  0.826543   0.579666   1.426   0.1540
## year_fact2020  1.123600   0.588754   1.908   0.0564 .
## year_fact2021  0.448981   0.617702   0.727   0.4673
## quarter_fact2 -0.038080   0.129896  -0.293   0.7694
## quarter_fact3  0.069978   0.132887   0.527   0.5985
## quarter_fact4 -0.288589   0.123832  -2.330   0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                    edf Ref.df      F  p-value
## s(catch)                         3.218  3.218 10.517 6.82e-07 ***
## te(ecd_lon,ecd_lat):year_fact2012 3.001  3.001  2.555   0.0536 .
## te(ecd_lon,ecd_lat):year_fact2013 3.000  3.000  1.293   0.2749
## te(ecd_lon,ecd_lat):year_fact2014 3.000  3.000  7.549 4.91e-05 ***
## te(ecd_lon,ecd_lat):year_fact2015 3.000  3.000  1.606   0.1879
## te(ecd_lon,ecd_lat):year_fact2016 3.000  3.000  1.354   0.2560
## te(ecd_lon,ecd_lat):year_fact2017 3.617  3.617  3.312   0.0163 *
## te(ecd_lon,ecd_lat):year_fact2018 7.533  7.533  7.678  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2019 3.000  3.000  2.512   0.0567 .
## te(ecd_lon,ecd_lat):year_fact2020 3.000  3.000  0.837   0.4752
## te(ecd_lon,ecd_lat):year_fact2021 3.000  3.000  7.716 4.04e-05 ***
## s(capacity)                      1.000  1.000  2.593   0.1074
## s(hours_since_sunrise)           1.000  1.000  0.122   0.7267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0439
##   Scale est. = 1         n = 5357
```

A QQ-plot for the GAM component of the GAMM model for presence-absence of silky shark can be found in Fig. 7. Marginal effects plots for smoothed predictors in the GAM model can be found in Fig. 8.

3.2.3 GAMM for log-abundance when present of silky shark

The ANOVA table for the LME model is:

```
##   numDF denDF  F-value p-value
## X    47  4329 63.96075  <.0001
```

The residuals versus fitted for the LME model can be found in Fig. 9.

The summary output for the GAM component of the GAMM model is as follows:

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## logsilky ~ s(catch, k = 7) + te(ecd_lon, ecd_lat, by = year_fact,
##     k = 7) + s(capacity, k = 6) + s(hours_since_sunrise, k = 6) +
##     follow + year_fact + quarter_fact
##
## Parametric coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.73836    0.23706   7.333 2.67e-13 ***
## followTRUE     0.04893    0.02785   1.757    0.079 .
## year_fact2013 -0.02495    0.30390  -0.082    0.935
## year_fact2014 -0.06676    0.23957  -0.279    0.781
```

```
## year_fact2015  0.00763    0.25025   0.030    0.976
## year_fact2016 -0.17902    0.23712  -0.755    0.450
## year_fact2017  0.20596    0.23669   0.870    0.384
## year_fact2018  0.00402    0.23589   0.017    0.986
## year_fact2019  0.11438    0.24080   0.475    0.635
## year_fact2020 -0.05200    0.24206  -0.215    0.830
## year_fact2021 -0.21793    0.25620  -0.851    0.395
## quarter_fact2 -0.05164    0.04708  -1.097    0.273
## quarter_fact3  0.03273    0.04768   0.686    0.492
## quarter_fact4  0.06315    0.04603   1.372    0.170
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                     edf Ref.df      F  p-value
## s(catch)                          3.699  3.699 44.464  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2012  4.381  4.381  4.376  0.00110 **
## te(ecd_lon,ecd_lat):year_fact2013  5.512  5.512  1.801  0.28708
## te(ecd_lon,ecd_lat):year_fact2014  3.000  3.000  3.838  0.00930 **
## te(ecd_lon,ecd_lat):year_fact2015 10.653 10.653  2.550  0.00379 **
## te(ecd_lon,ecd_lat):year_fact2016  3.003  3.003  2.583  0.05159 .
## te(ecd_lon,ecd_lat):year_fact2017  9.200  9.200  4.822 2.45e-06 ***
## te(ecd_lon,ecd_lat):year_fact2018  8.191  8.191  6.133  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2019 10.699 10.699  5.371  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2020  7.766  7.766 10.926  < 2e-16 ***
## te(ecd_lon,ecd_lat):year_fact2021  3.443  3.443 10.066 4.87e-06 ***
## s(capacity)                        1.000  1.000  0.750  0.38653
## s(hours_since_sunrise)             1.001  1.001  0.093  0.76077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.127
##   Scale est. = 0.785     n = 4387
```

A QQ-plot for the GAM component of the GAMM model for abundance when present of silky shark can be found in Fig. 10. Marginal effects plots for smoothed predictors in the GAM model can be found in Fig. 11.

3.3 Predicted CPUEs

We present here three sets of model predictions: (1) the predicted CPUE annual time series based on using the original data used to fit models as the prediction grid, presented purely as a check for the model's ability to reproduce the nominal CPUE given original inputs; (2) the standardized CPUE time series with target tuna catch per set allowed to vary across years; and (3) the standardized CPUE time series after removing interannual variability from tuna catch per set predictions, presented as a control for the potential impact of changing catchability over time that may not be fully accounted for by other predictors (e.g., `follow` and `hours_since_sunrise`) included in the models.

3.3.1 CPUE based on original data

The CPUE based on the original input data shown in Fig. 12 approximately agrees with the nominal CPUE, but there are nominal CPUEs outside the 95% prediction interval for the year 2012 and 2016. The excursion for 2016 is relatively small, whereas that for 2012 is more pronounced, suggesting that the model is not adequately reproducing catch patterns for this year with very low fishing effort and observer coverage of fishing effort (Fig. 2).

3.3.2 CPUE based on prediction grid including interannual variability in tuna catch per set

The standardized CPUE values shown in Fig. 13 show a strong temporal trend with nearly a two-fold increase in silky shark catch per set between the beginning and the end of the time series. Increases are concentrated in two specific periods 2012-2013 and 2016-2017.

3.3.3 CPUE based on prediction grid removing interannual variability in tuna catch per set

Removing interannual variability in target tuna catch produces a highly variable, but more of less horizontal temporal trend in catch per set (compare Fig. 13 with Fig. 14). The overall curve is generally similar to that of the nominal CPUE.

4 Discussion

The main standardized CPUE time series for silky shark (Fig. 13) indicates nearly a factor of two increase in silky shark abundance over the time period of this study. There are a number of arguments both for and against placing faith in this result. One argument for believing this trend is that the transition to low- or non-entangling dFADs that has occurred over the last decade could have significantly reduced previous-reported high rates of ghost mortality of silky sharks in dFADs (Filmalter et al. 2013). If so, it is plausible that this led to a rebound of silky shark populations. There is also anecdotal evidence from scientific tagging programs that it has become considerably easier to encounter silky sharks in PS catch, though this is not reflected in nominal CPUE values. Nevertheless, the period 2012-2014 during which time a lot of the overall increase occurred is likely too early for low-entangling dFADs to have had much impact on silky shark populations.

Arguments against believing this trend include the fact that much of the increase occurs during time periods that are special for one reason or another. The observer coverage is low for the 2012-2013 time period and this time period was also heavily perturbed by the threat of Somali piracy, both potentially biasing results. This possibility is supported by the fact that nominal CPUEs for 2012 do not agree with model predicted CPUEs predicted using the same data that was used to construct models (Fig. 12). 2012-2013 was also characterized by the transition to using echosounder buoys, which could impact both target and non-target catches (Wain et al. 2021), though it seems unlikely that this explains the majority of the increase in standardized CPUEs as Wain et al. (2021) found that echosounder buoys explained only about a 10% change in target tuna catch and we included the `follow` predictor variable in all models (though it was not significant in the target tuna model, suggesting that the effect of this parameter is being absorbed by other parameters in the model). Removing interannual variability from target tuna catch predictions did remove the upward trend in silky shark catch per set (Fig. 14), but it is difficult to imagine that all interannual change in catch per set is driven by changes in catchability (that we would want to remove), particularly as standardized skipjack catch per set values show an increasing trend in abundance over the ~2012-2020 time period (Kaplan et al. 2023).

The period 2016-2017 also saw a large increase in standardized CPUE values. 2017 was the onset of the quota for yellowfin tuna catch, which has significantly altered purse seine fishing strategies in the Indian Ocean. These changes could alter silky shark catch rates. Nevertheless, the principal impacts of the quota have been to reduce the frequency of fishing sets on free-swimming schools and increase the frequency of sets on FOBs, including a greater proportion of fishing sets on FOBs belonging to the fishing vessel (Floch et al. 2021, Wain et al. 2021; Kaplan et al., unpublished results). Within FOB fishing activities themselves, there has been relatively little change in fishing practice other than increased reliance on FOBs released by the fishing vessel. One would expect most of these changes to lead to a decrease in silky shark catch per set, though the timescales for such a decrease may be longer than a single year.

There are a number of aspects of the model that could potentially be improved. One is that the model does not include space-quarter interaction terms, potentially missing seasonal variability and explaining the difference between the nominal CPUE and the model predicted CPUE seen in Fig. 12. This limitation is conceptually easy to fix, but including all possible interactions makes models unstable and extremely long to run. Additional reflection on the appropriate way to address these issues is required.

Another concern of the approach is that target tuna catch is included as predictor in models of silky shark catch along with other predictor variables that we know impact target tuna catch, leading to a lack of independence between predictors. The solution to this issue is to use residuals instead of raw catch values after removing the impact of those other predictor variables. This solution was attempted and it seems unlikely that this would have a large impact on results as models based on residuals showed little difference to those based on raw catch, but future work should give additional consideration of this question.

A further improvement would be to use random forest models instead of GAMMs for the binomial silky shark presence-absence model as was done in Dumont et al. (submitted). Random forest models have the advantage of being inherently non-linear and highly flexible and have been shown repeatedly to be good models for binomial processes (e.g., Dumont et al. submitted).

Finally, there are some potentially interesting predictor variables that have not been included in models. Primary among these are dFAD densities. The main reason these were not included was due to the time that would be required to coordinate with the non-French components of the purse seine fleet to obtain dFAD densities. Nevertheless, this issue could be relatively easily addressed in future work.

5 Data availability statement

Given the confidential nature of the data used in this article, requests for data access should be addressed directly to the IRD's Exploited Tropical Pelagic Ecosystems Observatory (Ob7; https://www.ob7.ird.fr/en/) using the following e-mail address: adm-dblp@ird.fr.

Acknowledgements

We thank ORTHONGEL for making industry collected observer data via the OCUP observer program available for scientific analyses. We thank the IRD-Ob7 pelagic observatory of the MARBEC laboratory for tropical tuna logbook and observer data management and preparation, particularly L. Floch, P. Sabarros and A. Duparc.

Funding statement

This work was funded by the Ob7 Exploited Tropical Pelagic Ecosystems Observatory of the IRD via support from the the European Data Collection Framework (DCF, Reg 199/2008 and 665/2008) and the French "Direction des Pêches Maritimes et de l'Aquaculture" (DPMA), as well as by the Horizon Europe project REDUCE (101135583). Data collection was financially supported by EU via the Data Collection Framework (DCF, Reg 199/2008 and 665/2008), and by ORTHONGEL in the framework of their OCUP program.

References

Amandè M, Chassot E, Chavance P, Murua H, Delgado de Molina A, Bez N (2012) Precision in bycatch estimates : The case of tuna purse-seine fisheries in the Indian Ocean. *ICES Journal of Marine Science* **69**:1501–1510. doi:10.1093/icesjms/fss106

Andersen MM (2022) Prediction intervals for Generalized Additive Models (GAMs). *Mikkel Meyer Andersen*

Dumont A, Duparc A, Sabarros PS, Kaplan DM (submitted) Modeling bycatch abundance in tropical tuna purse seine fisheries on floating objects using the Δ method.

Fasiolo M, Nedellec R, Goude Y, Wood SN (2018) Scalable visualisation methods for modern generalized additive models. *Arxiv preprint*

Filmalter JD, Capello M, Deneubourg J-L, Cowley PD, Dagorn L (2013) Looking behind the curtain: Quantifying massive shark mortality in fish aggregating devices. *Frontiers in Ecology and the Environment* **11**:291–296. doi:10.1890/130045

Floch L, Marsac F, Fily T, Depetris M, Duparc A, Kaplan DM, Lebranchu J (2021) Statistics of the French purse seine fishing fleet targeting tropical tuna in the Indian Ocean (1981-2020). IOTC-2021-WPDCS17-21. Indian Ocean Tuna Commission Working Party on Data Collection and Statistics (WPDCS), Online/Virtual

Forget F, Muir J, Hutchinson M, Itano D, Sancristobal I, Leroy B, Filmalter J, Martinez U, Holland K, Restrepo V, Dagorn L (2021) Quantifying the accuracy of shark bycatch estimations in tuna purse seine fisheries. *Ocean & Coastal Management* **210**:105637. doi:10.1016/j.ocecoaman.2021.105637

Forget FG, Capello M, Filmalter JD, Govinden R, Soria M, Cowley PD, Dagorn L (2015) Behaviour and vulnerability of target and non-target species at drifting fish aggregating devices (FADs) in the tropical tuna purse seine fishery determined by acoustic telemetry. *Canadian Journal of Fisheries and Aquatic Sciences* **72**:1398–1405. doi:10.1139/cjfas-2014-0458

Goujon M, Maufroy A, Relot-Stirnemann A, Moëc E, Bach P, Cauquil P, Sebarros P (2017) Collecting data on board French and Italian tropical tuna purse seiners with common observers: Results of Orthongel's voluntary observer program OCUP (2013-2017) in the Indian Ocean. IOTC-2017-WPDCS13-22_Rev1. 13th IOTC Working Party on Data Collection and Statistics (WPDCS), Mahé, Seychelles

IOTC Secretariat (2022a) Executive Summary: Silky Shark. Appendix 21. Indian Ocean Tuna Commission

IOTC Secretariat (2022b) Updates on the implementation of the IOTC Regional Observer Scheme and its pilot project. IOTC-2022-SC25-07. 25th IOTC Sientific Committee, Victoria, Seychelles

Kaplan DM, Chassot E, Amandé JM, Dueri S, Demarcq H, Dagorn L, Fonteneau A (2014) Spatial management of Indian Ocean tropical tuna fisheries: Potential and perspectives. *ICES Journal of Marine Science* **71**:1728–1749. doi:10.1093/icesjms/fst233

Kaplan DM, Grande M, Alonso MLR, Duparc A, Imzilen T, Floch L, Santiago J (2023) CPUE standardization for skipjack tuna (Katsuwonus pelamis) of the EU purse-seine fishery on floating objects (FOB) in the Indian Ocean. IOTC-2023-WPTT25(DP)-11_Rev1. IOTC 25th Working Party on Tropcial Tunas, Data Preparatory Meeting, Online

Mannocci L, Forget F, Tolotti MT, Bach P, Bez N, Demarcq H, Kaplan D, Sabarros P, Simier M, Capello M, Dagorn L (2020) Predicting bycatch hotspots in tropical tuna purse seine fisheries at the basin scale. *Global Ecology and Conservation* **24**:e01393. doi:10.1016/j.gecco.2020.e01393

Pianet R, Pallares P, Petit C (2000) New Sampling and Data Processing Strategy for Estimating the Composition of Catches by Species and Sizes in the European Purse Seine Tropical Tuna Fisheries. WPDCS00-10. IOTC Proceedings no. 3

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

Rice J, Harley S (2013) Updated stock assessment of silky sharks in the western and central Pacific Ocean. *Scientific Committee Ninth Regular Session*:6–14

Tennekes M (2018) tmap: Thematic maps in R. *Journal of Statistical Software* **84**:1–39. doi:10.18637/jss.v084.i06

Torres-Irineo E, Amandè MJ, Gaertner D, Molina AD de, Murua H, Chavance P, Ariz J, Ruiz J, Lezama-Ochoa N (2014) Bycatch species composition over time by tuna purse-seine fishery in the eastern tropical Atlantic Ocean. *Biodiversity and Conservation* **23**:1157–1173. doi:10.1007/s10531-014-0655-0

Wain G, Guéry L, Kaplan DM, Gaertner D (2021) Quantifying the increase in fishing efficiency due to the use of drifting FADs equipped with echosounders in tropical tuna purse seine fisheries. *ICES Journal of Marine Science* **78**:235–245. doi:10.1093/icesjms/fsaa216

Wickham H (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Available from: https://ggplot2.tidyverse.org

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. *Journal of Open Source Software* **4**:1686. doi:10.21105/joss.01686

Wood SN (2017) Generalized additive models: An introduction with r, 2nd edn. Chapman; Hall/CRC

List of Tables

Table 1: Basic statistics regarding our catch and bycatch dataset from the French Indian Ocean purse seine fleet.

Statistic	Value
Time period	2012 - 2021
Total FOB sets	15,967.0
FOB sets with observers	5,357.0
Coverage of observer data (%)	33.6
Obs. sets with silky sharks	4,387.0
Prevalence of silky sharks (%)	81.9
No. silky shark individuals	40,022.0
Mean no. silky per set	7.0

Table 2: Description of predictors used in models of bycatch per FOB fishing set. The "Abbreviation" column gives the short name for each variable used in the models themselves. The "95% interval" and "Mean" columns refer to the central 95% of the data and the mean of the data over FOB fishing sets in our dataset.

Full name	Abbreviation	Units	Type	Median	95% interval
Lat. / Lon. recorded in logbook	ecd_lat / ecd_lon	Decimal degrees	Continuous	-	-
Quarter	quarter	[1:4]	Categorical	-	-
Year	year	[2012:2021]	Continuous	-	-
Trop. tuna catch	catch	Tonnes	Continuous	19.70	[2.02; 105.22]
Vessel capacity	capacity	m^3	Continuous	1519.05	[1206.00; 2119.00]
Set time relative to local sunrise	hours_since_sunrise	Decimal hours	Continuous	3.14	[-0.31; 10.28]
Fishing own FOB	follow	[TRUE,FALSE]	Boolean	-	-
Vessel identifier	vessel	-	Categorical	-	-

List of Figures

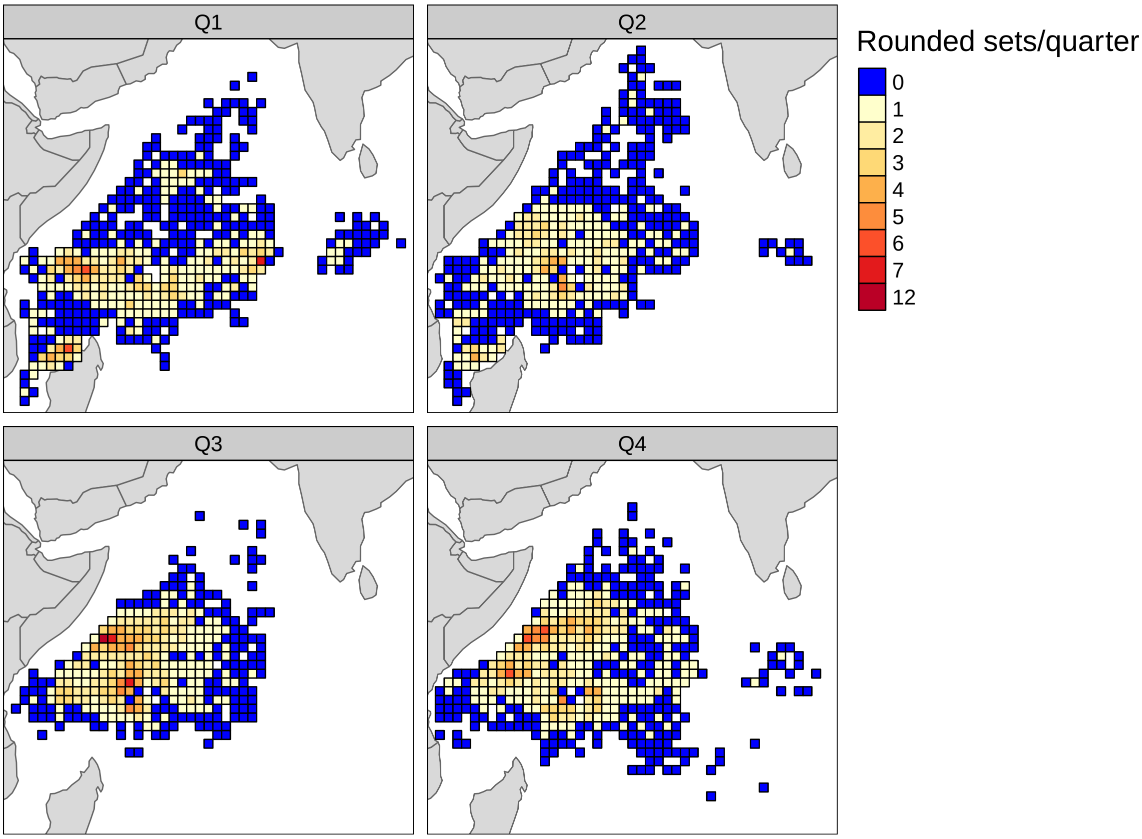

Figure 1: The rounded average (over years) number of FIB fishing sets per quarter in each 1° × 1° grid cell. Zeros not included in our final prediction grid for CPUE standardization are shown in blue, all other values are indicated by the graduated yellow-orange-red color scale.

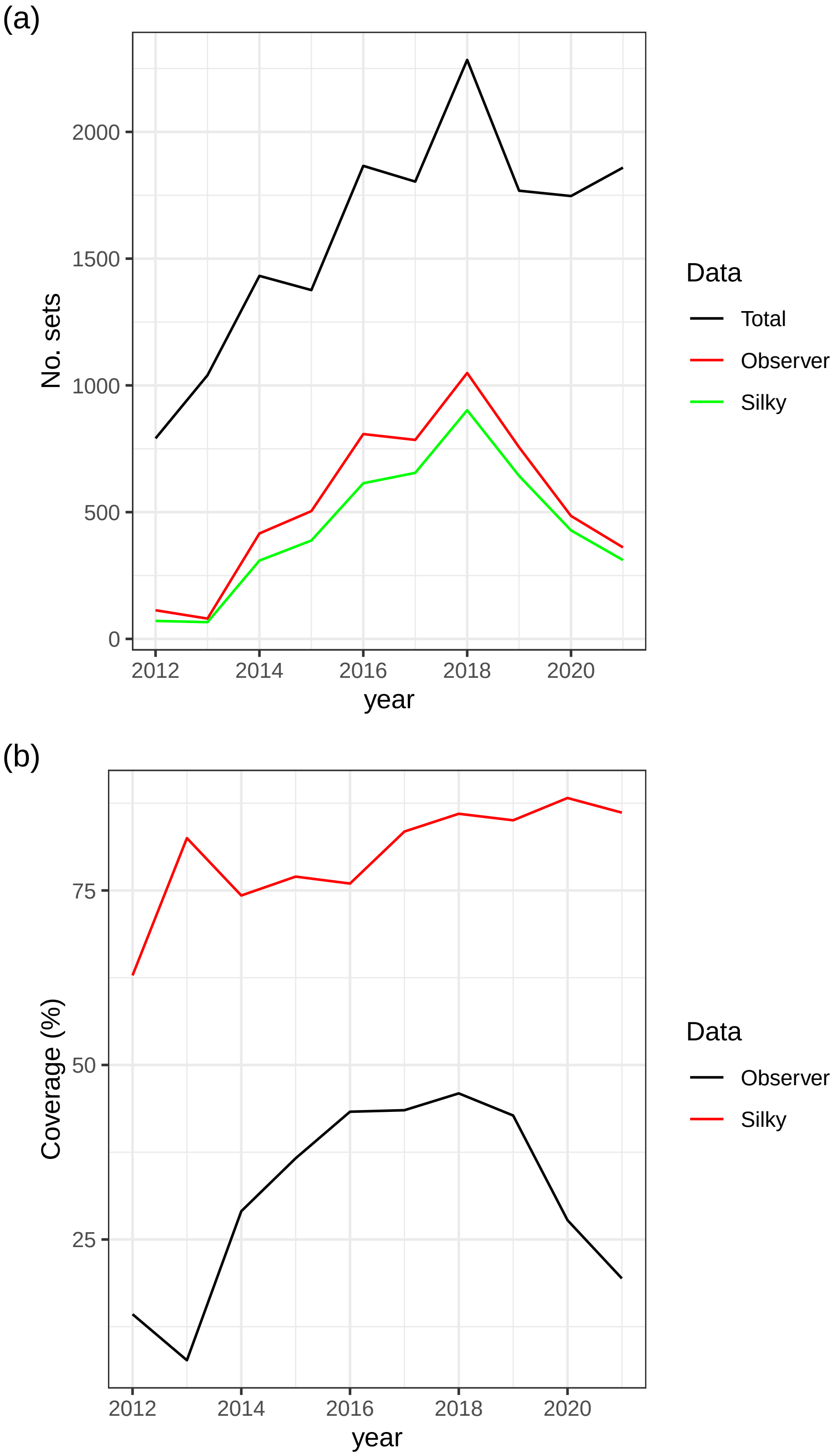

Figure 2: The temporal evolution across years of fishing effort and observer coverage. In (a), the black curve indicate the total number of fishing sets in a year, the red curve indicates the fishing sets with observer coverage and the green curve indicates the number of fishing sets for which silky sharks were found to be present in the catch. In (b), observer coverage is shown in black as a percentage of all fishing sets in the data used for fitting models, while prevalence of silky sharks in sets is shown in red as a percentage of all fishing sets with observer data.

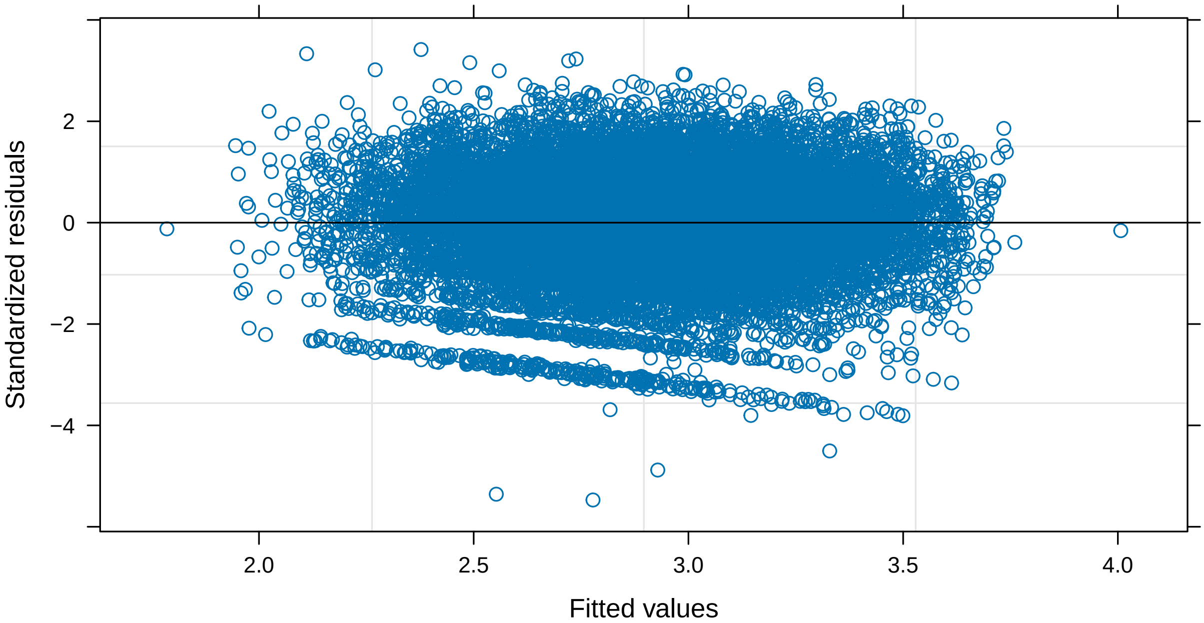

Figure 3: Standard plot of residuals versus fitted values for the LME component of the GAMM model for log-catch of target tunas.

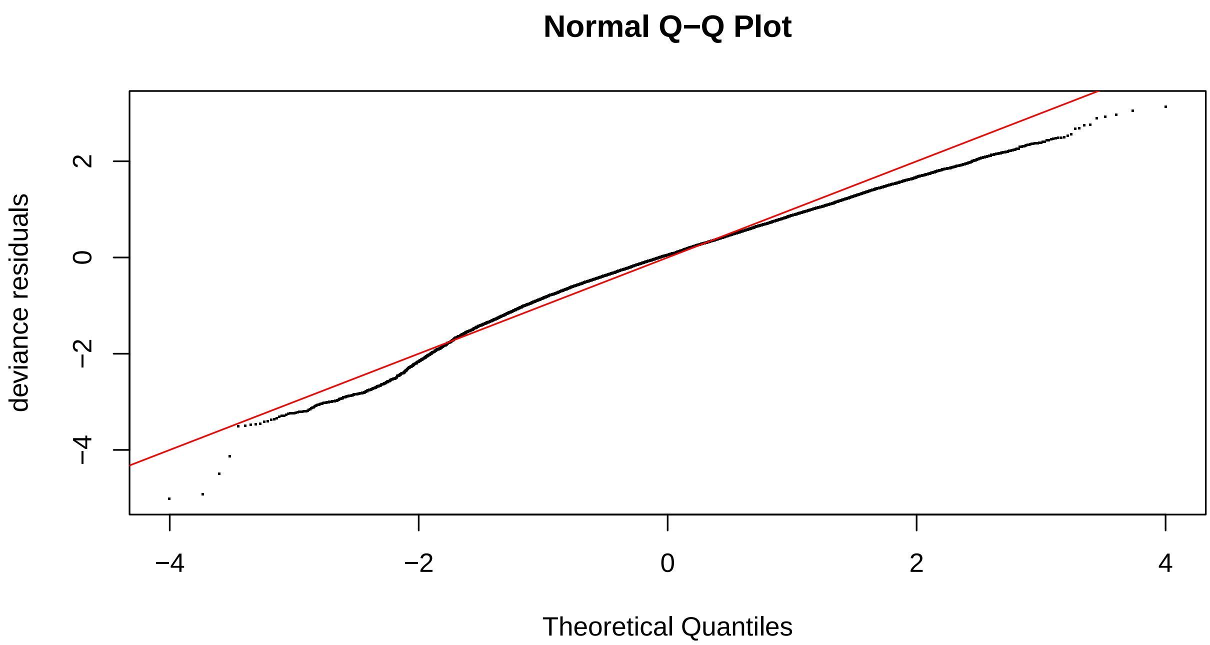

Figure 4: Standard QQ plot of predicted versus fitted values for the GAM component of the GAMM model for log-catch of target tunas.

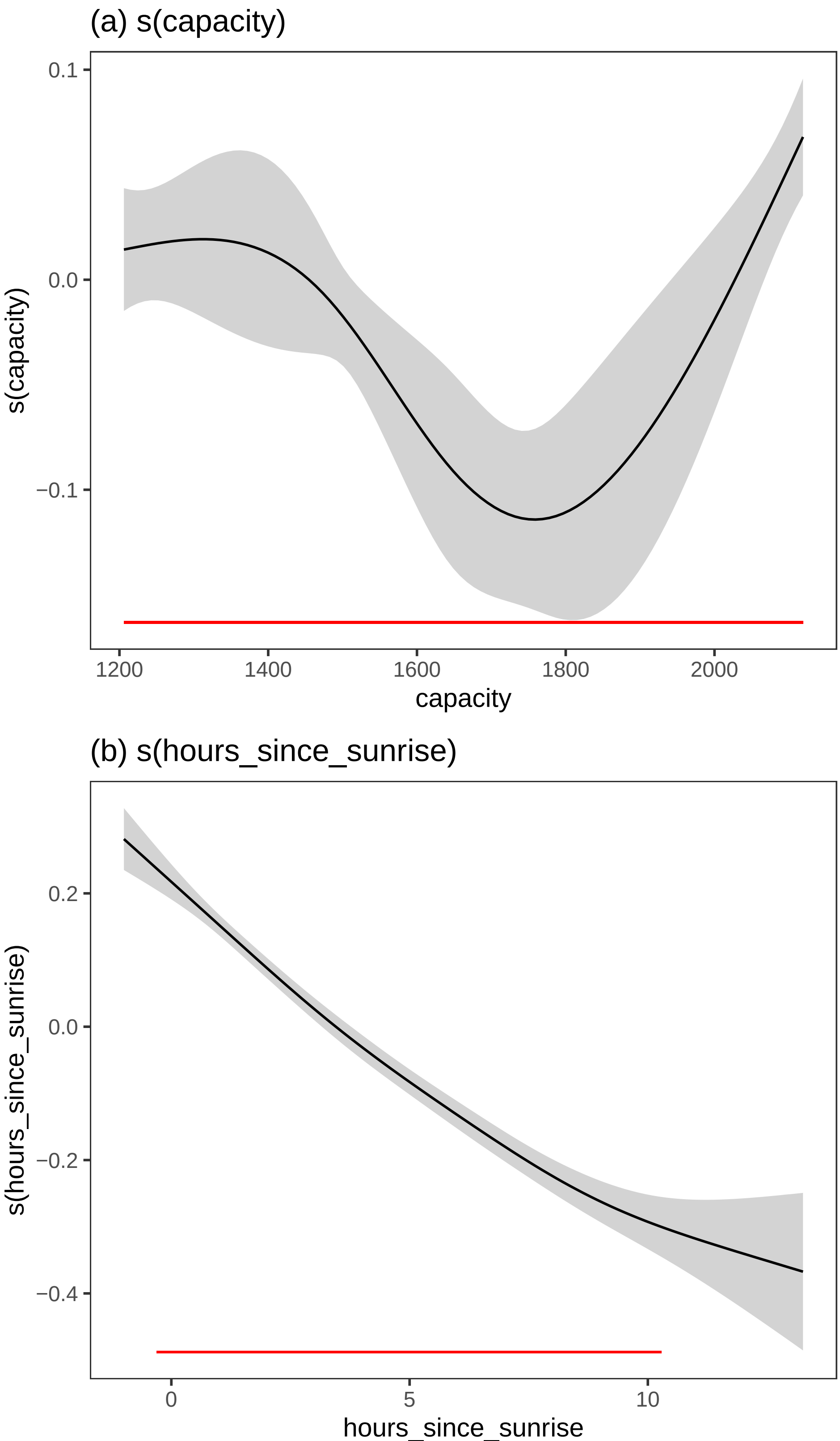

Figure 5: Marginal effects plots for the smooths on predictors `capacity` and `hours_since_sunrise` for the GAM component of the GAMM model for log-catch of target tunas. Red horizontal bars at the bottom of each figure indicate the central 95% of the data for the corresponding predictor.

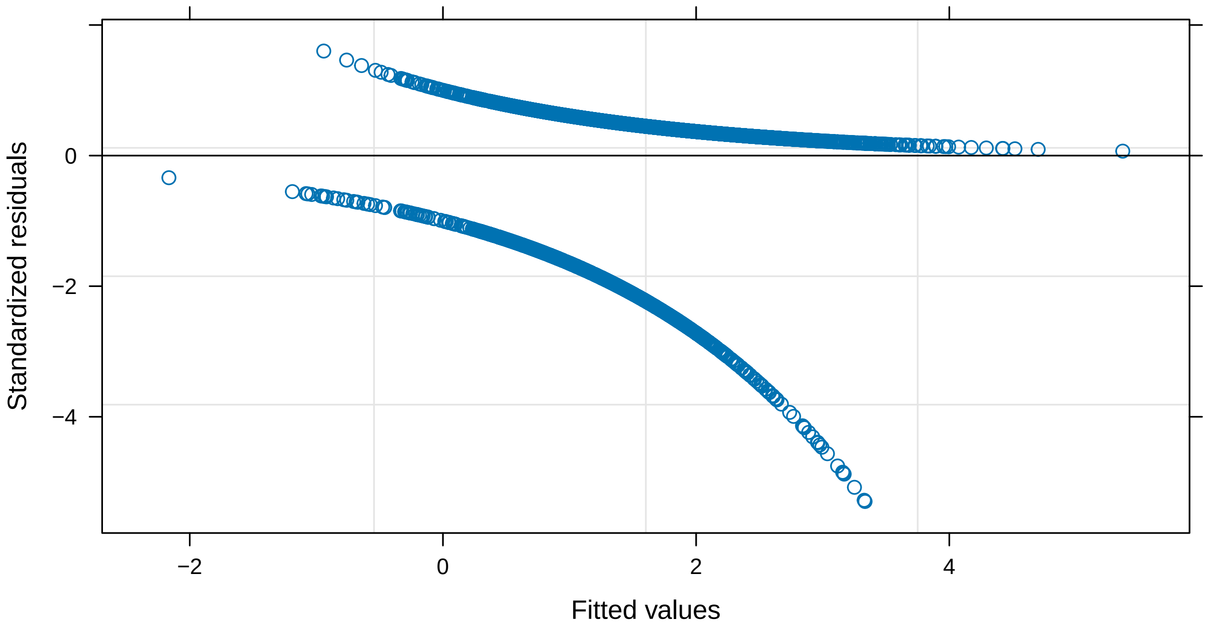

Figure 6: Standard plot of residuals versus fitted values for the LME component of the GAMM model for presence-absence of silky shark.

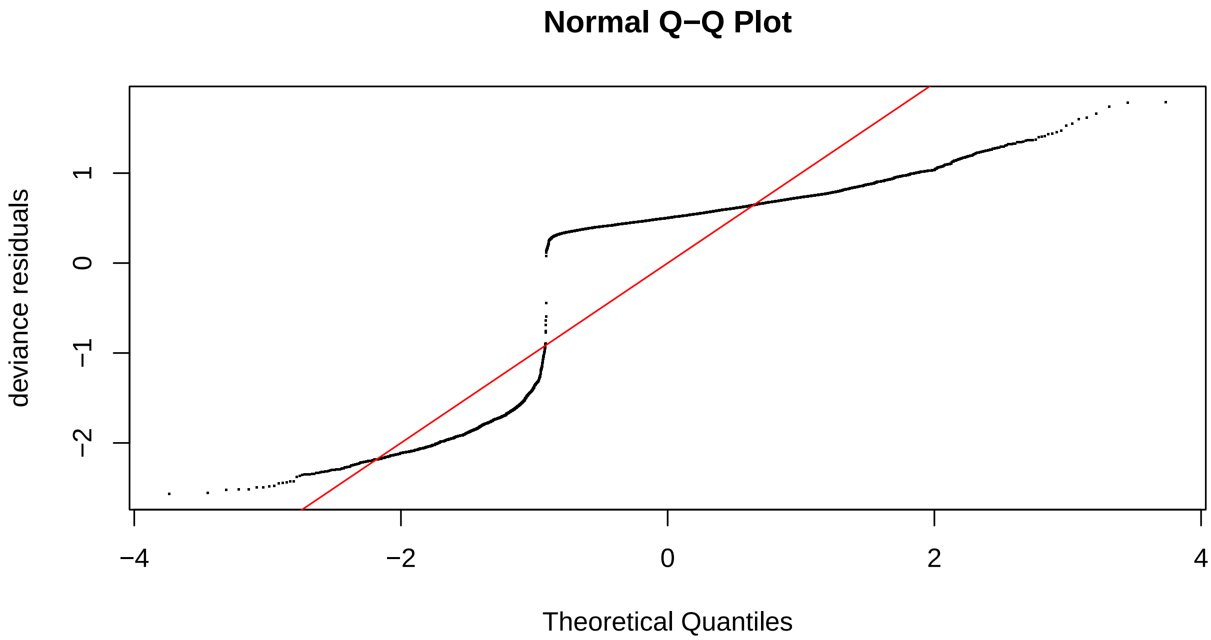

Figure 7: Standard QQ plot of predicted versus fitted values for the GAM component of the GAMM model for presence-absence of silky shark.

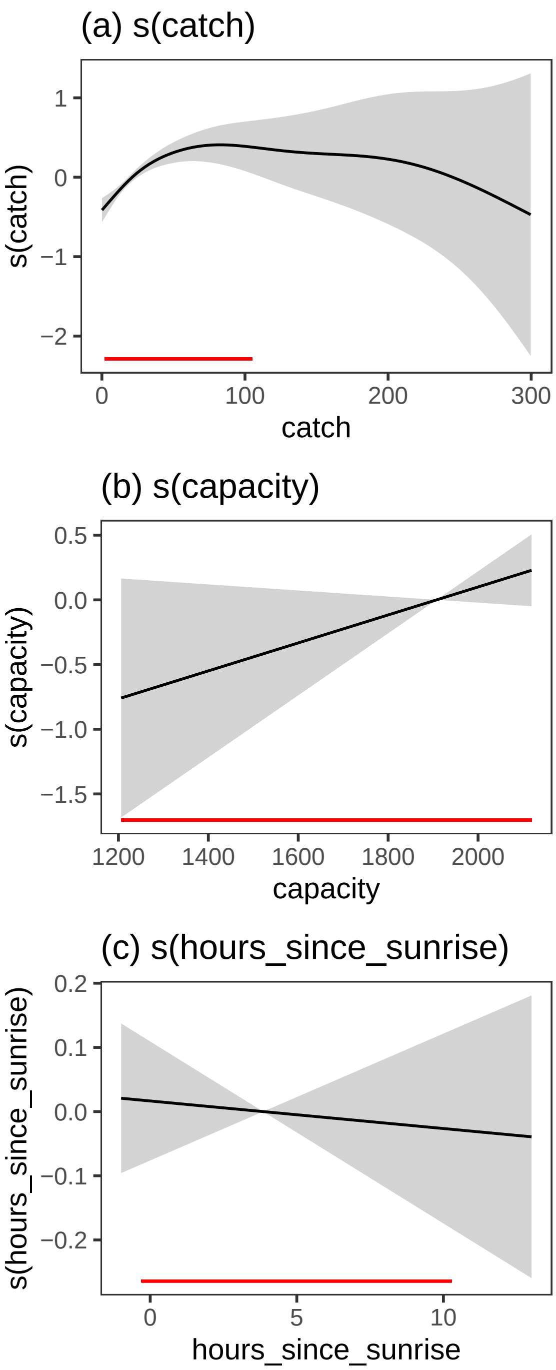

Figure 8: Marginal effects plots for the smooths on predictors `catch`, `capacity` and `hours_since_sunrise` for the GAM component of the GAMM model for presence-absence of silky shark. Red horizontal bars at the bottom of each figure indicate the central 95% of the data for the corresponding predictor.

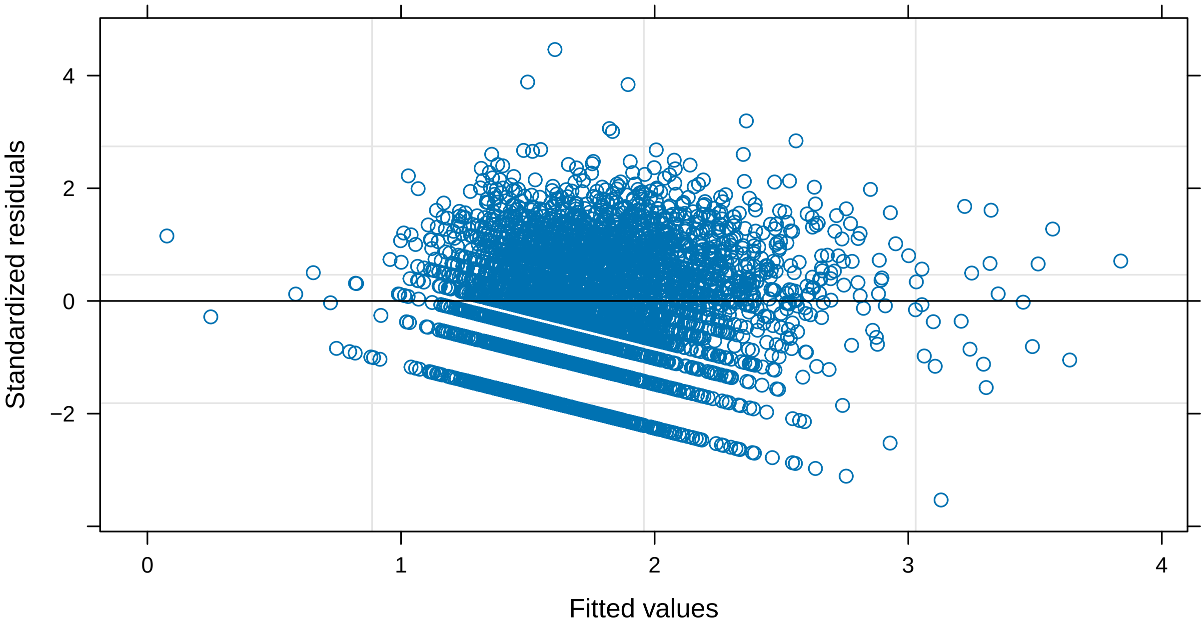

Figure 9: Standard plot of residuals versus fitted values for the LME component of the GAMM model for log-abundance when present of silky shark.

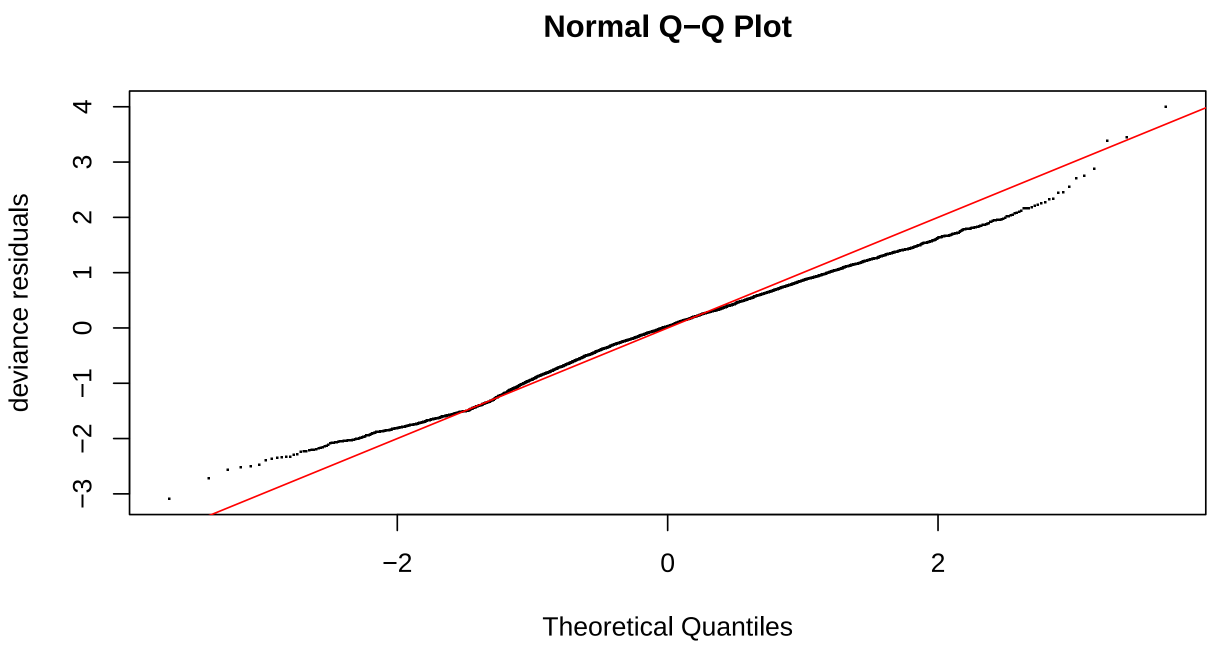

Figure 10: Standard QQ plot of predicted versus fitted values for the GAM component of the GAMM model for log-abundance when present of silky shark.

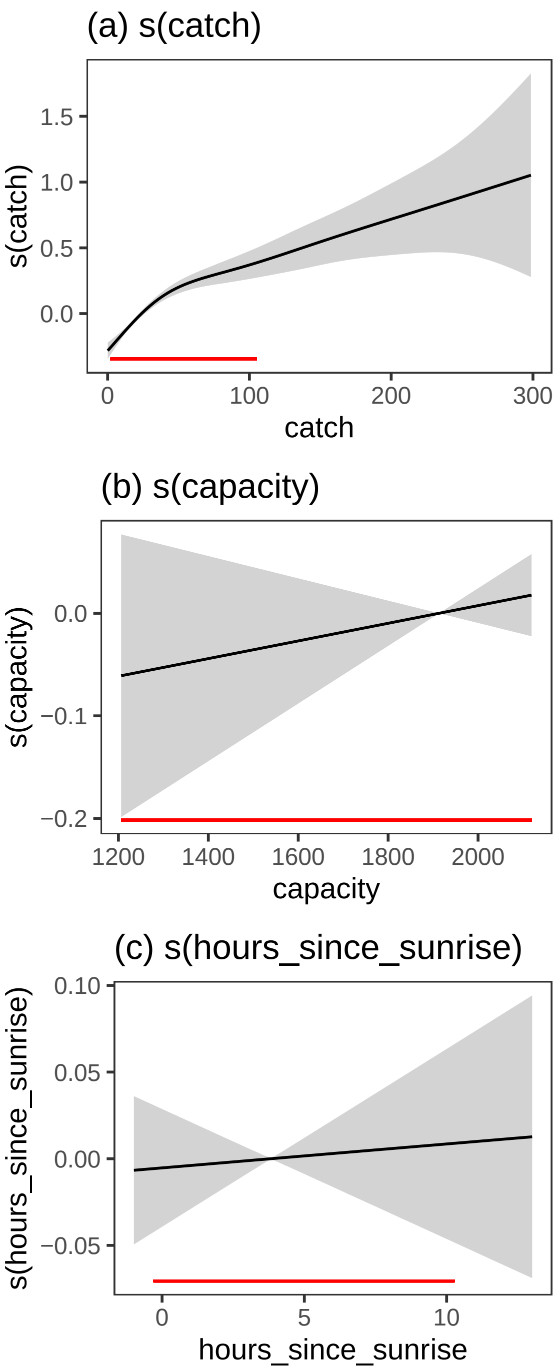

Figure 11: Marginal effects plots for the smooths on predictors `catch`, `capacity` and `hours_since_sunrise` for the GAM component of the GAMM model for abundance when present of silky shark. Red horizontal bars at the bottom of each figure indicate the central 95% of the data for the corresponding predictor.

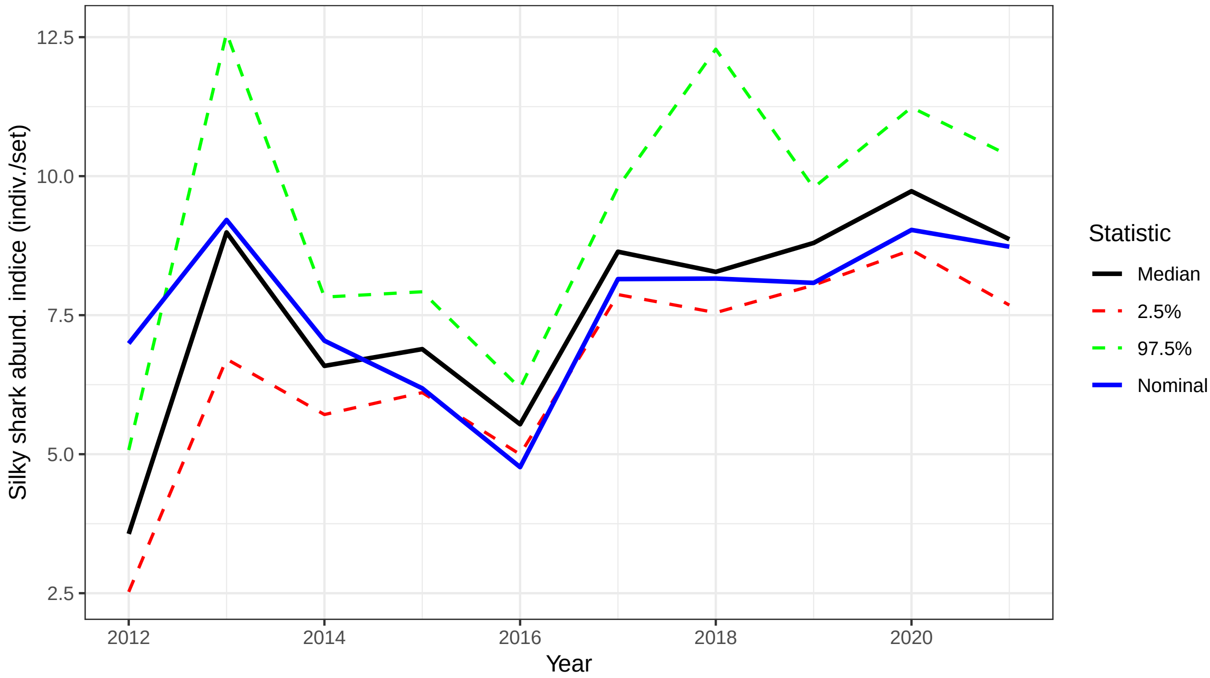

Figure 12: Silky shark estimated CPUE based on predictions of tuna catch per set. Predictions were carried out based on the original data used to train models so that nominal CPUE and predicted CPUE should be close.

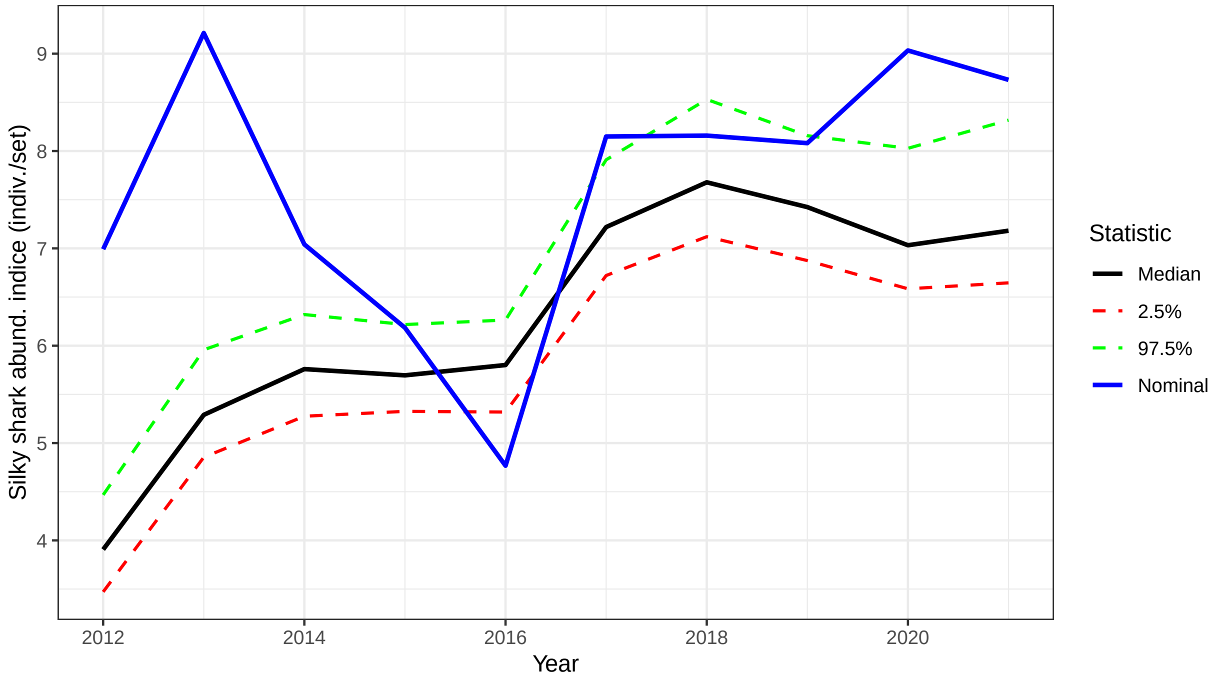

Figure 13: Silky shark standardized annual abundance indice based on prediction grid and predictions of tuna catch per set with interannual variability.

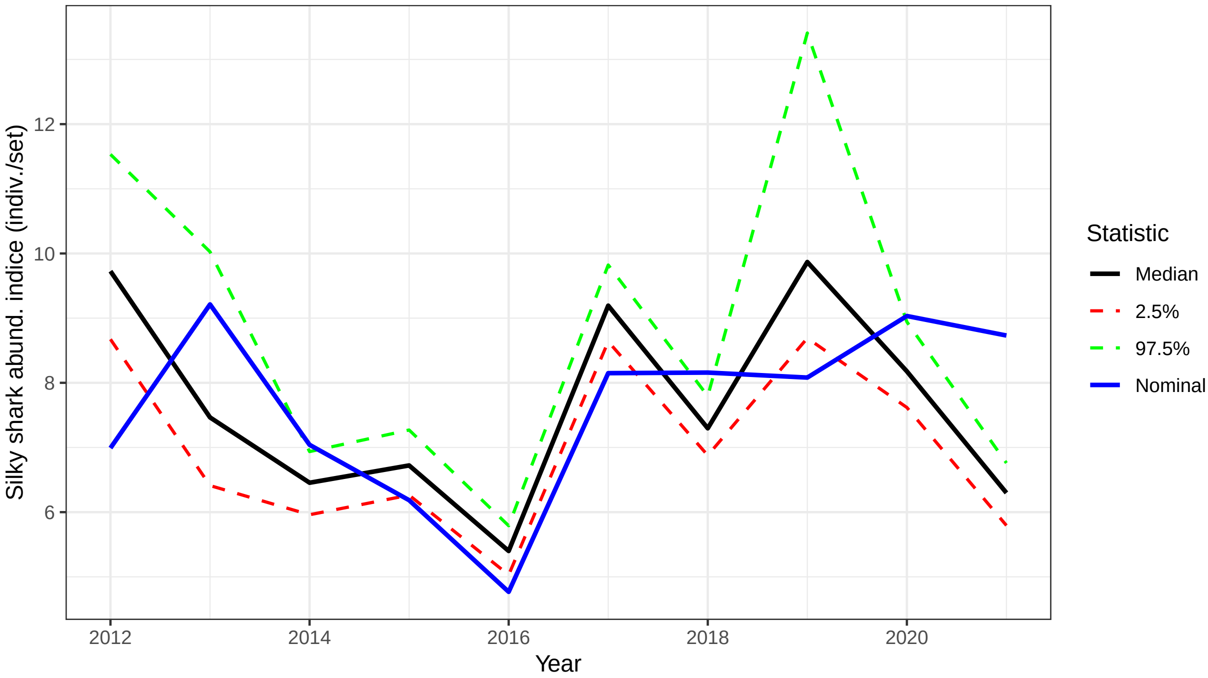

Figure 14: Silky shark standardized annual abundance indice based on predictions of tuna catch per set for which interannual variability has been removed (but quaterly and spatial variability remains).